15 Gender Differences in Stress, Trauma, and PTSD Research

Application of Two Quantitative Methods

LYNDA A. KING
HOLLY K. ORCUTT
DANIEL W. KING

When one reads or hears the phrase "gender differences," the statistical situation that most readily comes to mind is the simple comparison of the scores of women and men on a particular variable. For example, we might ask the research question: "Following exposure to a common stressor, do women and men differ in posttraumatic stress disorder (PTSD) symptomatology as measured by their continuous symptom severity scores on the PTSD Checklist (PCL; Blanchard, Jones-Alexander, Buckley, & Forneris, 1996; Weathers, Litz, Herman, Huska, & Keane, 1993)?" This comparison is typically accomplished in one of two ways. Probably the most direct avenue is to evaluate the significance of the difference between independent group means (the PCL mean for men vs. that for women) using a t test or its equivalent F ratio. This strategy allows us to determine whether, on average, men score higher than women, or women score higher than men on the PCL. Similar information is available by relying on a bivariate regression analysis in which the dependent variable of interest, PCL score, is regressed on the independent variable of gender, and gender is coded in a dichotomous fashion (such as 1 = women and 0 = men). With this 1/0 coding scheme, the intercept in the regression equation is the mean for men, and the regression coefficient is the difference between the means of women and men. The significance of the regression coefficient—the gender difference—again may be evaluated by t or F, and the strength of the relationship—between gender and PCL score—can be captured in a correlation coefficient. Either approach may be used, and in both cases, the interpretation of a significant finding is the same.

Another simple comparison between genders can be made when the outcome of interest is a frequency count, a percentage of occurrence, or some documentation of proportion within levels of a categorical variable. In this regard, it may be of interest to determine whether women and men have different prevalence rates of PTSD; that is, do the data suggest a tendency for the diagnosis of PTSD to be more (or less) prevalent in women (or men)? We would count the number of PTSD-positive and PTSD-negative diagnoses for women and men separately, and then compute a chi-square test of the association between gender and diagnosis. A significant value for the chi-square statistic in this instance would lead to the conclusion that there is an association between gender and diagnosis and, hence, that there is a difference in prevalence rates for women and men. This particular examination of gender differences in frequency counts, percentages, or proportions could be extended to include research situations in which the comparison of women and men involves distributions across more than two categories: for example, full PTSD, partial PTSD, and no PTSD.

For any of these gender difference analyses, it is also possible to introduce "control variables" or covariates that statistically equate study participants before performing the gender-based comparison. Such a strategy is aimed at ruling out alternative influences for an observed gender difference on the outcome variable and possibly providing a more sensitive or powerful test of the difference between women and men. Consider the contrast between women and men who have experienced a common traumatic event, such as a natural or human-made disaster in their community. If one were interested in gender differences in PTSD symptom severity or prevalence rates, it might be appropriate to equate statistically all disaster victims on the intensity of the exposure. Removing the influence of exposure intensity in the comparison between women and men would more clearly answer the question: "Given an equal degree of exposure to the event, do women and men differ with regard to PTSD?" If PTSD is measured as a continuous variable (e.g., a symptom severity score on the PCL), then the gender difference question could be addressed with an analysis of covariance, a randomized block analysis of variance, or a multiple regression analysis. If PTSD is assessed as a categorical variable (e.g., frequencies, percentages, or proportions of persons classified as PTSD-positive or PTSD-negative, or as full PTSD, partial PTSD, or no PTSD), then log linear and logistic regression analyses are available.

Finally, though not as frequently used by researchers in the realm of gender differences, there are statistical tests for the difference between group variances. These procedures enable the researcher to decide whether there is evidence for a difference in the dispersion or spread of scores on a continuous variable. Lunneborg and Abbott (1983) demonstrated the logic for a likelihood ratio-based chi-square test for the difference between group variances, and the Statistical Package for the Social Sciences (SPSS) software package provides Levene's test of homogeneity of variance, a non-likelihood, ratio-based test recommended for non-normal data. To illustrate, in the National Vietnam Veterans Readjustment Study (Kulka et al., 1990), the dispersion of scores on the Mississippi Scale for Combat-Related PTSD (Keane, Caddell, & Taylor, 1988) was greater for male veterans who served in Vietnam than for female veterans who served in that war. The standard deviation for men was 22.35, whereas the standard deviation for women was 17.21, and the computed Levene statistic is significant, $F(1, 1617) = 47.07$, $p < .001$. Therefore, the spread or variance of scores for men is of higher magnitude than the spread or variance of scores for women. This finding is not surprising given that the subsample of women, mostly nurses, was quite homogeneous in education and other background characteristics, and their war-zone exposures were more uniform than those of the men who served in Vietnam (King, King, Foy, & Gudanowski, 1996; King, King, Gudanowski, & Vreven, 1995).

Thus, there is an entire collection of well-used methods that can identify differences between genders in their group mean scores, in their respective frequencies of occurrence of some event or condition, or in the dispersion of their scores. Regardless of the complexity of the analyses, the "bottom line" is to compare women and men on some aspect of the variable of interest.

Yet another type of research question is very important for a comprehensive understanding of gender differences. The question goes beyond the simple issue of differences in means, frequencies, or variances, and emphasizes female–male differences in the associations or relationships between or among variables. In this chapter, this notion of differential relationships as a function of gender provides an integrating theme to describe two seemingly unrelated methods. We begin with a presentation of moderated multiple regression analysis in which the interest is whether the relationship between a predictor and an outcome (say, intensity of trauma exposure and PTSD, or attitudes toward women and perpetration of domestic violence) is different for women than for men. We then discuss multiple group confirmatory factor analysis, the aim of which is to determine if the relationship between a latent variable or factor (e.g., PTSD) and a measured or observed variable (e.g., an item score, or perhaps a composite score on a handful of items measuring the same PTSD symptom cluster) is different for women than for men. Throughout our presentation, we attempt to provide a simple, not too mathematical conceptualization of the methods. We supplement, where possible, with illustrations and our own analyses with existing data sets.

MODERATED MULTIPLE REGRESSION

Perhaps the simpler of our two approaches to examining differential relationships as a function of gender is moderated multiple regression. In this section, we briefly introduce multiple regression. Then, we describe the concepts of interactions and moderator variables, followed by the special circumstance in which gender moderates the relationship between two continuous variables. We next demonstrate a moderated multiple regression analysis of the interaction between scores on a prominent measure of gender role ideology and gender in the prediction of intimate relationship violence. The section concludes with brief commentary on other issues and concerns, and recommendations for further reading on the topic.

Overview of Multiple Regression

Multiple regression is a class of statistical procedures intended to describe the association between an outcome, or dependent variable, and a weighted composite of two or more predictors, or independent variables. A very simple representation of a multiple regression equation would be

$$Y_i = B_0 + B_1X_{1i} + B_2X_{2i} + \ldots + B_kX_{ki} + e_i \qquad \text{(Eq. 1)}$$

Y_i is the dependent variable for person i; X_{1i}, X_{2i}, and as many as X_{ki} are independent variables for person i. The multiple regression equation has its roots in the algebra of a straight line, in that B_0 is the Y-intercept and the B_1 through B_k values are slopes. Each of these slopes indicates the change in the dependent variable (Y) per unit change in its associated independent variable (e.g., X_1) at constant or specified values of the other independent variables ($X_2 \ldots X_k$). The slopes also denote the unique contribution of an independent variable and are referred to as "regression coefficients." The e_i symbol recognizes the presence of a residual, or error, in the prediction of any single person's score on the dependent variable from the weighted composite of the various independent variables.

The weights in the composite (B_0 and B_1, $B_2 \ldots B_k$) are derived so as to satisfy optimally some specified function or criterion. One such criterion is the minimization of error, such that across all cases in the sample, the sum of the squared error terms (values of e) is as small as it can possibly be. Weights derived in this way are called "ordinary least squares estimates" and are the most typically used weights in multiple regression analysis. Under certain assumptions, these weights are maximum likelihood estimates, which allow for the implementation of parametric statistical theory. In other words, tests of statistical significance can be employed for both the intercept and regression coefficients, and for the model as a whole.

The independent variables may be either categorical (e.g., victim of prior violence vs. not a victim of prior violence; male vs. female gender) or continuous (e.g., a scaled measure of the severity of trauma exposure; age at time of exposure). The dependent variable likewise may be either categorical (as in the diagnosis of PTSD as present or absent) or continuous (as in symptom severity scores on the PCL or Mississippi Scale). We are particularly interested in the application of multiple regression to the case in which the dependent variable is continuous. (The case in which the dependent variable is categorical is beyond the scope of this chapter, and we refer the reader to major texts by Agresti, 1990, and Hosmer & Lemeshow, 1989.) Furthermore, in pursuit of our goal of describing differential relationships as a function of gender, our independent variables will be a mixture of categorical and continuous, with gender serving in the special role of moderator.

Introduction to Interactions and Moderator Variables

An understanding of differential relationships between a predictor and an outcome as a function of gender is most easily represented in a traditional 2 × 2 factorial design, in which the two levels of the gender categorical variable are crossed with two levels of another categorical variable. Let us assume that this second variable, denoted gender role ideology, classifies individuals with regard to their basic attitudes about the equality of women and men as either traditional or egalitarian. The dependent variable is a self-report of the number of incidents of physical violence directed toward an intimate partner. In other words, the researcher is asking whether the propensity to commit violent acts is a function of the individual's gender (constant over, or regardless of, gender role ideology) and/or a function of the individual's traditional–egalitarian attitudes about the equality of women and men (constant over, or regardless of, her or his gender). Each of these research questions defines a main effect. But a more interesting question may be the extent to which these variables *jointly* influence intimate relationship violence, such that some unique combination of the levels of gender and gender role ideology serves either to exacerbate or mitigate incidents of violence. This latter question reflects the notion of an interaction.

This research situation with two sets of contrived data is depicted in Table 15.1. Table 15.1a shows a pattern of cell means (for our dependent variable, intimate relationship violence) that might be obtained when we find main effects for both gender and ideology, but no interaction. Table 15.1b shows an alternative pattern of cell means that would indicate an interaction between gender and ideology (and also a main effect for ideology, but none for gender). The marginal values for the rows, rightmost on both tables, are the means for each level of ideology averaged over the levels of gender, and the marginal values for the columns, at the bottom of

TABLE 15.1. Illustration of Main Effects and Interaction: A Simple 2 × 2 Factorial Design

a. Main effects: No interaction

		Gender F	Gender M	
Ideology	Traditional	1.50	2.50	2.00
	Egalitarian	1.00	2.00	1.50
		1.25	2.25	

b. Gender × gender role ideology interaction

		Gender F	Gender M	
Ideology	Traditional	1.00	2.50	1.75
	Egalitarian	2.00	.50	1.25
		1.50	1.50	

Note. The dependent variable is the average number of incidents of intimate relationship violence. F, female; M, male.

each table, are the means for each level of gender averaged over the levels of ideology.

Concentrating first on Table 15.1a, the difference between the row marginal values (2.00 – 1.50 = .50) is the same as the difference between the traditional and egalitarian levels of ideology for both women (1.50 – 1.00 = .50) and men (2.50 – 2.00 = .50). Similarly, the difference between the column marginal values (1.25 – 2.25 = –1.00) is the same as the difference between women and men for both the traditional (1.50 – 2.50 = –1.00) and egalitarian (1.00 – 2.00 = –1.00) levels of ideology. This pattern of fictitious findings contrasts with that depicted in Table 15.1b. The column marginal values in Table 15.1b (both 1.50) suggest that there is no gender main effect; on average, women and men are reporting similar amounts of violence against their intimate partners. There is an apparent main effect for ideology; those holding traditional attitudes report more violence (1.75) than those holding egalitarian attitudes (1.25). Going further, beginning with the rows, the difference between the marginal values (1.75 – 1.25 = .50) is not equivalent to either the difference for women only (1.00 – 2.00 = –1.00) or the difference for men only (2.50 – .50 = 2.00). Likewise, the difference between the column marginal values (1.50 – 1.50 = 0) is not equal to the corresponding differences for individuals with a traditional ideology (1.00 – 2.50 = –1.50) or for individuals with an egalitarian ideology (2.00 – .50 = 1.50). In the instance in which there is no interaction (Table 15.1a), the effect of a given variable (e.g., ideology) is constant across the levels of the other variable (e.g., gender). When there is an interaction (Table 15.1b), the effect of a given variable is not constant across the levels of the other variable.

In our Table 15.1b example, women who are traditional in their ideology report perpetrating fewer incidents of violence against their partners (mean of 1.00) than women who are egalitarian in their ideology (mean of 2.00), whereas the opposite pattern is seen for men, with traditional men reporting perpetration of more violence (mean of 2.50) than egalitarian men (mean of .50). Thus, the apparent influence of gender role ideology on intimate relationship violence differs as a function of gender; stated alternatively, the association between ideology and violence is different for women and men. We conclude that there is a gender × gender role ideology interaction effect in accounting for intimate relationship violence.

Figure 15.1 is intended to clarify the results presented in Table 15.1, especially the idea of an interaction in terms of differential relationships as a function of gender. In both Figures 15.1a and 15.1b, the relationship between ideology and violence is graphed separately for women and men by plotting values of the cell means (average violence scores) for each of the two levels of ideology (traditional and egalitarian).

In Figure 15.1a, where there is no interaction, the main effect for gender is captured in the observation that the line for men is higher on the violence scale than the line for women. Similarly, the main effect for ideology is captured in the observation that both lines are negatively inclined; hence, persons with traditional attitudes are scoring higher on the violence scale than persons with egalitarian attitudes. The parallelism, or uniform slopes, of the two lines denote the lack of an interaction: The change in violence scores, or difference in cell means, for women and men remains constant across both levels of the ideology variable.

In Figure 15.1b, the interaction is obvious. Here, the relationship between ideology and violence clearly differs for women and men. Note the slopes of the separate lines. For women, the change in violence scores, or difference in cell means, is positive when moving from traditional to egalitarian ideology; for men, the corresponding change is negative. Thus, in predicting the propensity to inflict intimate relationship violence from a person's gender role ideology, it is extremely important to know the gender of the person as well. Gender can be said to be a moderator of the relationship between ideology and violence. It might be noted that interactions can be interpreted bidirectionally. For this illustration, we could have pursued the relationship between gender and intimate relationship violence as a function of gender role ideology. But because this chapter focuses on methods of detecting gender differences, gender is the appropriate moderator variable.

Gender in Interaction with Continuous Independent Variables

Our example treated gender role ideology as a categorical variable having two levels. The informed reader of gender research, however, is aware that

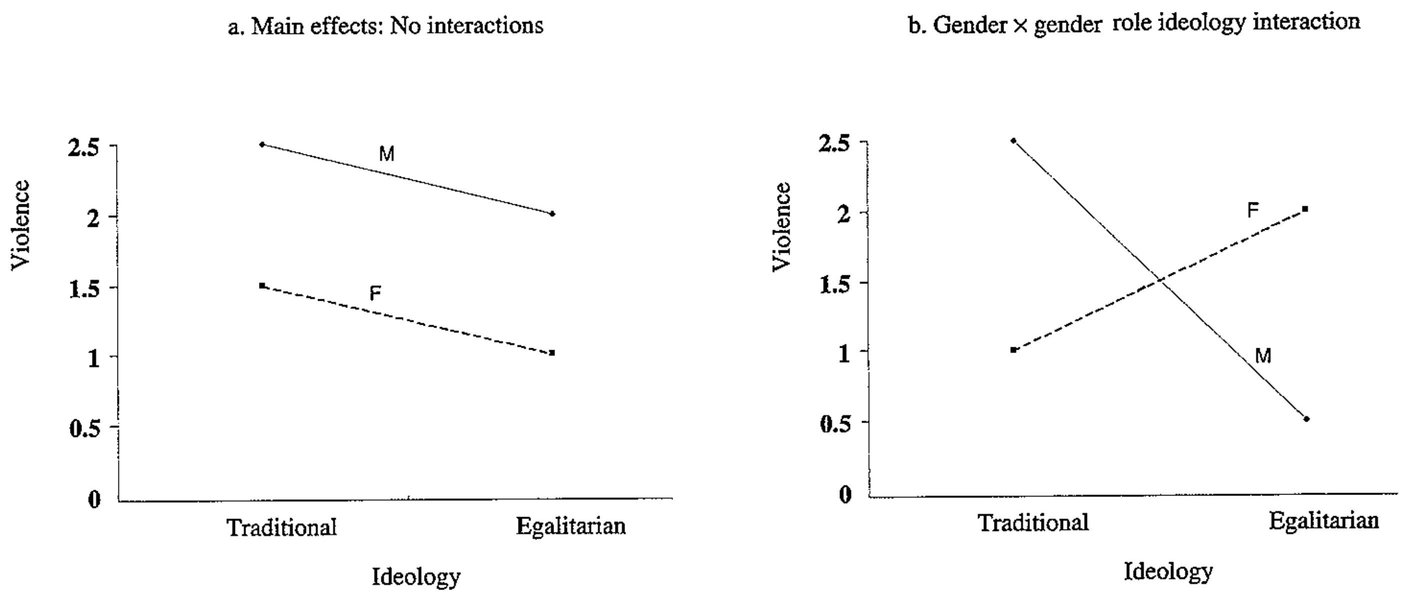

FIGURE 15.1. Graphical representation of the contrived data for the 2 × 2 factorial design. The dependent variable is the average number of incidents of intimate relationship violence. F, female; M, male.

this type of attitudinal variable is typically considered and measured as an ordered continuous dimension having multiple possible scores, ranging from low (traditional) to high (egalitarian); see, for example, the broadly used Attitudes Toward Women Scale (Spence & Helmreich, 1972), and the many assessment instruments catalogued by Beere (1990). Consequently, a more realistic representation of gender as a moderator of the ideology–violence relationship requires that the interaction represent the joint effect of the categorical gender variable and a continuous ideology variable. This generalizes to many questions in stress, trauma, and PTSD research, where the essential concern for female–male comparisons focuses on differential relationships between one or more continuous independent variables (e.g., intensity of trauma exposure, age at time of exposure, extent of social support network) and a continuous dependent variable (e.g., PTSD symptom severity). Thus, we extend our presentation to interactions in a multiple regression framework.

In multiple regression, an interaction between two predictor variables is represented by a third variable that is computed as the product of the two predictors. For example, a variable representing the interaction between X_1 and X_2 would be calculated as the product of scores on X_1 and X_2 for all cases; a variable representing the interaction of X_1 and X_3 would be calculated as the product of scores on X_1 and X_3 for all cases; and so on. With just two predictor variables and their interaction, the regression model would be

$$Y_i = B_0 + B_1X_{1i} + B_2X_{2i} + B_3X_{1i}X_{2i} + e_i \qquad \text{(Eq. 2)}$$

where all elements are as described previously (Eq. 1), the X_1X_2 product term carries the information of the interaction (Cohen & Cohen, 1983; Cohen, Cohen, Aiken, & West, 2002), and the B_3 regression coefficient delineates the unique contribution of the interaction term to the dependent variable. As noted by Jaccard, Turrisi, and Wan (1990), this coefficient is the change in relationship between one of the independent variables (say, X_2) and the dependent variable (Y) per unit change in the other independent variable (X_1), and vice versa. This interpretation can be seen by some minor manipulation of Eq. 2, most notably factoring X_2 from the third and fourth elements to the right of the equal sign:

$$Y_i = (B_0 + B_1X_{1i}) + (B_2 + B_3X_{1i})X_{2i} + e_i \qquad \text{(Eq. 3)}$$

Notice that this expression follows the general form for the equation of a straight line, documenting the relationship between two variables, X_2 and Y:

$$Y = (\text{intercept}) + (\text{slope})X_2 + \text{error} \qquad \text{(Eq. 4)}$$

Here, the slope for the regression of Y on X_2 is $(B_2 + B_3X_1)$ and the intercept can be reconceptualized as $(B_0 + B_1X_1)$. Thus, the relationship between Y and X_2, as represented by the slope, changes as X_1 takes on different values. And, for each unit change in X_1, the slope for the regression of Y on X_2 changes by B_3 units, the value of the regression coefficient for the interaction term. By designating X_1 as gender, we now have a mechanism for representing differential relationships between the other independent variable (X_2) and the dependent variable (Y). The model can be expanded to accommodate more independent variables, either categorical or continuous, some possibly intended also to interact with gender, with each other, or not. The works by Cohen and Cohen (1983) and Cohen et al. (2002) provide excellent detailed elaboration on more complicated multiple regression analyses. What follows is a presentation of the analysis and interpretation of findings for a set of real data, incorporating additional basic concepts germane to the method.

Demonstration of Moderated Multiple Regression Analysis Using Existing Data

Description of the Data

For demonstration purposes, we draw from a portion of a study of intimate relationship violence among college students (Fitzpatrick, Salgado, Suvak, King, & King, 2002). The data set comprises the responses of 219 college students (155 women and 64 men) to three measures: (1) the physical violence subscale of the Conflict Tactics Scale (CTS; Straus, 1979), (2) the Dyadic Adjustment Scale (DAS, Spanier, 1976), and (3) the Sex-Role Egalitarianism Scale—Form KK (SRES; King & King, 1993, 1997). The CTS physical violence subscale is the measure of intimate relationship violence, with higher scores indicating more violent acts perpetrated by the study participant on her or his partner within a 1-year time frame. The DAS indexes the quality of the participant's relationship with her or his partner, again with higher scores representing more satisfaction and comfort with the relationship. The SRES assesses gender role ideology and provides continuous scores along a dimension bounded by traditional (lower end) to egalitarian (upper end). Lower scores reflect beliefs in line with more conventional role expectations for women and men, whereas higher scores indicate beliefs about the behaviors and characteristics of women and men that are more flexible and less tied to conventional role expectations. The primary research question of interest is as follows: Controlling for relationship quality, does the association between gender role ideology and intimate relationship violence differ as a function of participants' gender? Thus, this is a modest extension of the fictitious data previously used as a vehicle in this chapter.

Analyses

The gender variable was dummy coded, assigning a value of 1 to the female participants and a value of 0 to the male participants. In actuality, any two numbers could have been used, but this 1/0 coding scheme is recommended to simplify computations and afford direct interpretation of regression coefficients in terms of gender differences. Also recommended is the practice of centering any variable that is expected to interact with gender (e.g., Aiken & West, 1991; Jaccard et al., 1990; Cohen et al., 2002), in this case, gender role ideology as measured by the SRES. Centering is quite simple. It is transforming the distribution of original or observed scores, so that the mean of the transformed scores will be 0. This was easily accomplished by subtracting the original SRES mean from every original SRES score. The interaction term was next computed by multiplying the dummy coded gender score by the centered SRES score. The distribution of DAS scores was also centered to simplify later computations.

Using a standard statistical software package, a two-step hierarchical multiple regression procedure (Cohen & Cohen, 1983) was performed. At the first step, CTS scores were regressed on centered DAS scores, centered SRES scores, and gender. At the second step, CTS scores were regressed on centered DAS scores, centered SRES scores, gender, and the interaction between gender and SRES (ideology).

Results and Their Interpretation

Table 15.2 presents the results of the hierarchical multiple regression analysis predicting intimate relationship violence (CTS violence subscale scores). Note that this table provides only unstandardized regression coefficients (*B*s); standardized coefficients (βs; betas) are omitted, because they are not as useful in interpreting interactions in moderated multiple regression. As shown in Table 15.2, at Step 1, relationship quality as measured by the DAS is the only significant variable. As one might predict, higher levels of relationship quality are associated with lower amounts of intimate relationship violence. The weighted composite of the three independent variables (DAS, SRES, gender) yielded a multiple correlation coefficient (R) of .201; 4% (the square of .201) of the variance in intimate relationship violence was accounted for by this composite.

When the interaction term (gender × SRES) was entered into the regression equation at Step 2, the value of R increased to .287, indicating that just over 8% (the square of .287) of the variance in intimate relationship violence was accounted for by the weighted composite (DAS, SRES, gender, gender × SRES). The significance of this increment (from 4% to 8%) is represented either in terms of a significant F statistic for the R^2 change, provided by many statistical software packages, or easily calculated (Cohen &

TABLE 15.2. Hierarchical Multiple Regression Results

Step	Variables in equation	Unstandardized coefficient (B)	Standard error	t	R
1	Intercept	1.837	.519	3.542***	
	DAS	–.049	.020	–2.395*	
	SRES	–.030	.021	–1.430	
	Gender	–.363	.621	–.585	.201
2	Intercept	1.415	.526	2.690***	
	DAS	–.046	.020	–2.306*	
	SRES	–.119	.035	–3.382**	
	Gender	–.030	.618	–.048	
	Gender × SRES	.135	.043	3.115**	.287

Note. DAS, Dyadic Adjustment Scale; SRES, Sex-Role Egalitarianism Scale.
*$p < .05$; **$p < .01$; ***$p < .001$.

Cohen, 1983; Cohen et al., 2002). The same information is carried in the significant t statistic (3.115) accompanying the regression coefficient for the interaction term, displayed in Table 15.2. Thus, there is support for the assertion of an interaction between gender and gender role ideology.

Note that, at Step 2, the unique contribution of SRES was significant, whereas at Step 1, it was not. An explanation for this shift can be found in the interpretation of the regression coefficients when interactions are not present versus when they are present. At Step 1, with no interaction being evaluated, the regression coefficient for SRES documents the effect of gender role ideology, holding constant gender and relationship quality. At Step 2, with the introduction of the interaction term, the SRES variable is evaluated for the condition in which the gender variable has a value of 0. In particular, its regression coefficient registers the association between gender role ideology and intimate relationship violence for male participants in the study, controlling for DAS. In this example, the regression coefficient is negative (–.119), indicating that as SRES scores increase toward the egalitarian end of the dimension, men tend to commit fewer acts of violence toward their partners. Conversely, the nonsignificant gender effect at Step 2 tells us that there is likely no difference between women and men in the amount of violence they perpetrate at an SRES value of 0, the mean SRES score following centering, again controlling for DAS. The intercept at Step 2 (1.415) denotes the score on the CTS violence subscale for men (coded 0) with an average score on the SRES (0) and on the DAS (0).

Using the regression coefficients from Step 2 of the multiple regression results, we can develop the equation for the regression of intimate relationship violence on relationship quality, gender role ideology, gender, and the interaction between gender and ideology. For simplicity, we drop the *i* subscript but retain the e component in recognition of error in prediction.

$$Y = 1.415 - .046X_D - .119X_S - .030X_G + .135X_GX_S + e \qquad \text{(Eq. 5)}$$

where Y is the predicted CTS score for person i, X_D is the DAS score, X_S is the SRES score, X_G is gender, and X_GX_S is the interaction term for person i. We may use this equation to address further our key research question regarding differential relationships between intimate relationship violence (Y) and gender role ideology (X_S) as a function of gender (X_G). Arranging terms in a manner similar to that depicted in Equations 3 and 4, we have

$$Y = (1.415 - .046X_D - .030X_G) + (-.119 + .135X_G)X_S + e \qquad \text{(Eq. 6)}$$

Substituting the mean DAS value (0), and alternative values for gender (1 and 0), we can develop two equations, one for the regression of violence on ideology for women, and one for men. For women, our equation becomes

$$Y = (1.415 - .046{*}0 - .030{*}1) + (-.119 + .135{*}1)X_S + e \qquad \text{(Eq. 7)}$$

$$Y = 1.385 + .016X_S + e \qquad \text{(Eq. 8)}$$

For men, the equation is

$$Y = (1.415 - .046{*}0 - .030{*}0) + (-.119 + .135{*}0)X_S + e \qquad \text{(Eq. 9)}$$

$$Y = 1.415 - .119X_S + e \qquad \text{(Eq. 10)}$$

Observe that the intercept and slope for men recapitulates the interpretations stated previously.

The two regression lines are graphed as Figure 15.2. It appears that the relationship between ideology and violence for women is only very slightly positive (regression coefficient of .016), whereas the relationship for men tends to be negative and stronger (regression coefficient of –.119). The significant regression coefficient for the gender × ideology interaction (Table 15.2) tells us that these relationships differ as a function of gender, and the value of this regression coefficient (.135) is the actual difference between the separate regression coefficients [.016 – (–.119) = .135]. But we can go further yet and determine the significance of the relationship within each gender, which is very likely of much interest to gender differences researchers.

Jaccard et al. (1990), among others, have provided a mechanism for evaluating the significance of the relationship between two continuous variables within each group defining the levels of a moderator variable. This

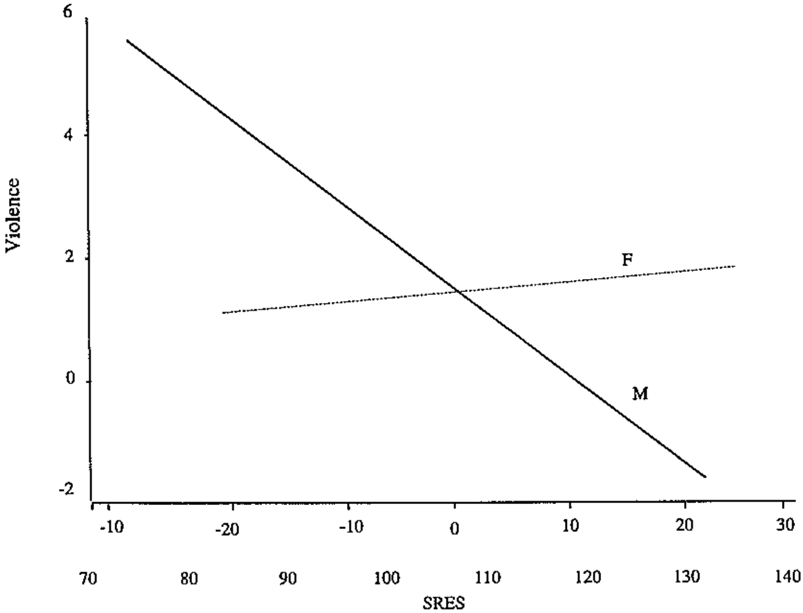

FIGURE 15.2. Association between gender role ideology and intimate relationship violence for women and men. Upper: centered scores; lower: raw scores.

procedure involves the computation of the standard error of the regression coefficient within a group, derived as a function of the variances and covariances among the components used to calculate that regression coefficient. To illustrate, in creating the regression equation for women in our sample (Eq. 7 and Eq. 8), two elements produced the final regression coefficient, –.119 and .135. These are the regression coefficients for SRES and gender × SRES taken from Table 15.2. Using the variances and covariance of these two elements—available as a part of multiple regression programs within most statistical software packages—standard errors can be calculated (see Jaccard et al., 1990, Appendix A), and tests of significance can be performed. For our data, the ideology–violence relationship was nonsignificant for women, $t(214) = .64$, NS, and significant for men, $t(214) = -3.38$, $p < .01$. An alternative method for directly computing and evaluating individual group simple slopes is provided by Cohen et al. (2002).

Additional Comments and Recommended Resources

Our presentation of moderated multiple regression necessarily is introductory and therefore limited and incomplete. Beyond the analyses already demonstrated, more questions can be answered about the nature of an interaction involving gender as a moderator variable. For example, as mentioned earlier, centering a predictor and dummy coding gender enables one to interpret the significance of the gender main effect at the mean of the predictor. But one can also go further and determine the significance of the difference between women and men on the outcome at any value of the predictor. In addition, in some circumstances, the researcher might wish to incorporate a curvilinear relationship between predictor and outcome, and propose that this relationship differs for women and men. The moderated multiple regression procedure allows for power polynomials and other strategies for assessing such curvilinearity. A three-way interaction between gender and two continuous predictors is possible, yielding information regarding differences in the interaction of the two continuous variables as a function of gender. Moreover, other categorical variables can be crossed with gender—race or ethnic identity and sexual orientation come to mind—and appropriately coded and included in the analysis to address more sophisticated questions about gender interactions. A set of multiple predictors also may be simultaneously incorporated in interaction with gender.

There are a few caveats in the search for gender differences in relationships using this methodology. Foremost is that interactions are often difficult to detect and replicate (Zedeck, 1971). What is more, when significant, the effect sizes are typically quite small, accounting for as little as 1–3% of the variance in the dependent variable (Champoux & Peters, 1987; McClelland & Judd, 1993). Interaction terms are known to be less reliable than their components and yield biased underestimates of their regression parameters. Nonetheless, moderated multiple regression, retaining the full range of scores on a predictor variable, is judged far superior to adopting a simple 2×2 factorial configuration, where scores on the predictor are dichotomized at some cutpoint (Cohen, 1983; West, Aiken, & Krull, 1996).

A few recommendations for further reading include the classic work by Cohen and Cohen (1983) and its successor (Cohen et al., 2002), and the texts by Darlington (1990), Draper and Smith (1998), and Pedhazur (1997), which provide comprehensive treatments of multiple regression and its applications, including interactions. The monographs by Jaccard et al. (1990) and Aiken and West (1991) focus exclusively on the analysis and interpretation of interactions and expand and supplement the issues presented here. There are many articles in the scientific literature on interactions and moderator variables. Among them are two of special note given the context of this section of the chapter: McClelland and Judd (1993) detailed reasons for failure to find or replicate interaction effects in field research, and Jaccard and Wan (1995) followed with an excellent assessment of problems with the reliability of interaction effects, closing with a suggestion for latent variable analysis (a topic to be introduced in the next section of this chapter). We also recommend the article by West et al. (1996), a demonstration of the analysis of interactions between categorical and continuous variables in the context of personality research; its content can easily be extrapolated to research on stress, trauma, and PTSD. Finally, O'Connor (1998) offers useful and free SAS and SPSS scripts to aid in the graphing of interactions. His programming was used to graph the interaction for this moderated multiple regression (Figure 15.2).

MULTIPLE-GROUP CONFIRMATORY FACTOR ANALYSIS

In our experience, PTSD theoreticians, researchers, and clinicians often pose the question of whether PTSD as a clinical entity or psychological variable is really the same for women as it is for men. This question of comparability of PTSD over genders addresses the issue of construct equivalence as represented in the literature on factorial invariance in confirmatory factor analysis (e.g., Byrne, Shavelson, & Muthen, 1989; Hancock, Stapleton, & Berkovits, 1999; Little, 1997; McArdle & Cattell, 1994; Meredith, 1993). In effect, the examination of factorial invariance can be couched in terms of the general theme of this chapter: here, the extent to which the relationship between a psychological construct, factor, or latent variable (PTSD) and scores on a manifest indicator of that variable (how an individual answers a PTSD item or set of items), differs for women compared to men. In this section, we concentrate on the confirmatory factor-analytic aspect of structural equation modeling. First, we provide a brief conceptual introduction to confirmatory factor analysis. Second, we discuss multiple-group modeling in confirmatory factor analysis and place the topic of factorial invariance in the context of gender differences. Third, we use an existing data set from a large-scale study of female and male Gulf War veterans to demonstrate a multiple-group confirmatory factor analysis. As before, we conclude with some commentary on issues and concerns, and recommendations for further reading.

Overview of Confirmatory Factor Analysis

Confirmatory factor analysis is a subset of the class of procedures known as structural equation modeling (e.g., Bollen, 1989; Bollen & Long, 1993; Hayduk, 1987, 1996; Hoyle, 1995; Kline, 1998; Loehlin, 1998; Marcoulides & Schumacker, 1996; Schumacker & Lomax, 1996; Schumacker & Marcoulides, 1998). It is the methodology associated with the specification and evaluation of the measurement component of a proposed structural equation model. This measurement component portrays hypothetical constructs or latent variables in terms of their observed scores or manifest indicators, typically, responses to items on some psychometric instrument.

Confirmatory factor analysis is used to test hypotheses about the underlying structure of the data; in other words, to what extent are there common entities or factors responsible for relationships or covariance patterns among the observed scores? The researcher must propose a priori a structure based on theory and/or previous factor analyses (exploratory or confirmatory), and then test how well the proposed structure fits the observed data.

To place confirmatory factor analysis in the context of multiple regression, as discussed in the previous section, we can express an individual's observed score on a particular PTSD item as follows:

$$Y_{ij} = B_{Y0} + B_{Y1}F_{i1} + B_{Y2}F_{i2} + \ldots + B_{Yk}F_{ik} + r_{ij} \quad \text{(Eq. 11)}$$

In this equation, the dependent variable (Y_{ij}) represents a score for person i on item j. The independent variables (F_{i1}, F_{i2} . . . F_{ik}) are scores on factors or latent variables for person i. These scores are not observed, and the number of factors is indeterminate and must be designated in advance by the researcher. As in the multiple regression equation described in the first section of this chapter, B_{Y0} is the Y-intercept, and B_{Y1}, B_{Y2} . . . B_{Yk} are slopes or regression coefficients. In this case, they are called "factor loadings" and denote the contribution of each factor or latent variable to the prediction of the observed score. The more a factor is implicated in an observed score, the stronger will be the factor loading or regression coefficient. The residual, a composite of both error and specificity carried by the observed score or manifest indicator, is symbolized as r_{ij}. For a given confirmatory factor analysis, there are as many equations as there are items or manifest indicators (Weathers, Keane, King, & King, 1997).

In proposing a factor structure in a confirmatory factor analysis, the researcher stipulates whether the intercepts, regression coefficients (factor loadings), and residuals of these equations, as well as factor means and factor variances, are to be estimated, or whether they should take on a specified value, usually 0 or 1. Such decisions essentially define the hypothesized pattern of loadings, the number of factors, which factors are predictors of which manifest indicators, and characteristics of the latent variable(s). One or more of these elements also may be subject to equality constraints, wherein they are specified to be equivalent to other elements, or some function of other elements. Additional considerations in model specification include whether the factors will be allowed to associate with one another, and whether residuals will be free to covary. Under most conditions, each manifest indicator is specified to load on only one factor; its loading on the other factors are fixed at 0 to facilitate interpretation of the solution. Also, residuals are normally not free to covary in confirmatory factor analysis. As a consequence of these decisions, a conceptual "mapping" of the postulated factor structure results, including a factor loading matrix augmented by intercepts for the regression of manifest indicators on factor scores, a matrix of residual variances for the manifest indicators, a matrix of variances and covariances among the factors, and a vector of factor means.

A number of methods for estimating the parameters in the hypothesized factor structure are available. Each seeks to optimize the "fit" between relationships actually observed in the data (typically characteristics of the manifest indicators: their means, variances, and covariances) and those that can be reproduced from the parameter estimates (intercepts, loadings, variances, etc.) in the hypothesized factor structure. Under certain assumptions, and for designated estimation procedures, the resulting fit function, a weighted sum of squared deviations multiplied by the number of cases minus 1, yields a noncentral chi-square statistic. Its degrees of freedom are equal to the difference between the number of means, variances, and unique covariances among the manifest indicators, and the number of parameters estimated in the model. An estimate of the noncentrality parameter is equal to the value of the computed noncentral chi-square minus the number of degrees of freedom. (The noncentrality parameter and the degrees of freedom are required to define the noncentral chi-square distribution.)

Low values of chi-square, relative to the degrees of freedom, indicate good fit and thus support the hypothesized factor structure. High values of chi-square, relative to the degrees of freedom, cast doubt upon the researcher's a priori factor structure. Given inherent problems with the chi-square statistic as the sole index of model–data fit, several additional goodness-of-fit indices have been developed (for profiles of those available, see Hu & Bentler, 1995, 1998; Joreskog & Sorbom, 1993a; Marsh, Balla, & McDonald, 1988). The more prominent fit indices at this writing include the Tucker–Lewis index (TLI; Bentler & Bonett, 1980; Tucker & Lewis, 1973), the comparative fit index (CFI; Bentler, 1990), the LISREL goodness-of-fit index (GFI; Joreskog & Sorbom), the root mean square error of approximation (RMSEA; Steiger, 1990), and the standardized root mean square residual (SRMR; Hu & Bentler, 1998). As we demonstrate later, there are procedures for evaluating the best among a series of competing models.

Multiple-Group Modeling: The Search for Construct Equivalence

Within the realm of structural equation modeling—and particularly for confirmatory factor analysis—we have the capability to pose questions about the similarity or equivalence of networks of relationships across two or more mutually exclusive groups. When multiple-group modeling is applied to confirmatory factor analysis, we are asking the extent to which a postulated factor structure holds across groups. In the earlier section on moderated multiple regression, we emphasized differential relationships between predictors and outcomes as a function of gender. In multiple-group confirmatory factor analysis, the representation is parallel, with manifest indicators being outcomes or dependent variables, and the factors being predictors or independent variables. There are slopes and intercepts, and they may or may not differ across genders. For example, when women and men respond to items on the PCL in reference to a common traumatic event (e.g., a natural disaster within their community, exposure to political terror and violence in their homeland, service in a war zone), we may be interested in whether the 17 items, scored for severity, have the same meaning for women and men. Stated another way, to what degree do the same factors appear relevant across genders and seem to contribute equally to reports of symptom severity for women and men? Other questions are also possible, the important point being that multiple-group confirmatory factor analysis supplies a mechanism for assessing comparability of factor structures for women and men (or any other grouping typology of interest).

The process of determining group differences in factor structures involves evaluation of a sequence of nested factor models. One model is said to be nested within another model when the pattern of associations present in the former is the same as for the latter, but there are fewer parameter estimates in the former than in the latter. A factor-analytic model having a larger number of parameter estimates is said to be more saturated than a model with a smaller number of parameter estimates. Other conditions held constant, the larger the number of parameter estimates in a model, the smaller will be the value of the chi-square statistic, which, as noted earlier, is an index of the discrepancy between the means, variances, and unique covariances of the observed data, and those reproduced by the proposed model. Put another way, the more information in the proposed model—a larger number of parameters being estimated with fewer degrees of freedom—the closer will be the model–data fit. The goal of science, of course, is parsimony, the quest to explain observed behavior or any other phenomenon as simply as possible (Mulaik & James, 1995). Therefore, for any structural equation modeling endeavor, we desire to represent our observations with a model containing a minimum of superfluous, irrelevant, or unnecessary parameters. As we evaluate or compare a series of nested models, we are searching for the simplest, least parameterized model that also retains strong model–data fit.

With regard to differences between factor structures for women and men, the test of nested models primarily emphasizes the notion of equivalence: Are the parameters that describe the factor structure for women equivalent to the parameters that describe the factor structure for men? Which are equivalent, and which are not equivalent? If two parameters are set equal to one another in a structural equation model, only one parameter is actually being estimated, not two. A multigroup confirmatory factor-analytic model with equality constraints on, for example, the factor loadings across groups, has fewer parameter estimates than one with factor loadings estimated separately for each group. Hence, a model with equality constraints is nested within a model without such constraints. By systematically imposing equality constraints of parameters across genders, we are able to compare more constrained with more saturated models and evaluate equivalence, thereby appraising gender differences.

A number of statistical methods are available for comparisons among constrained models. A frequently used approach is the test of significance of the difference between the chi-square statistics for two nested models. Keeping in mind that the chi-square statistic for a more constrained model is likely greater than (and never less than) that for a more saturated model, and that the chi-square statistic is an indicator of "misfit," the concern is whether the imposition of an equality constraint damages (as evidenced by an increase in chi-square) the model–data fit sufficiently to declare that the parameter(s) should not be constrained to be equivalent. Alternatively, is the damage so minor that a more parsimonious model (with equivalent values for the parameters across groups) is a better representation of the data? Sequential chi-square difference testing (Anderson & Gerbing, 1988; Steiger, Shapiro, & Browne, 1985), with the difference between two chi-square statistics evaluated at the difference between their associated degrees of freedom, assists in making these judgments. It has been suggested (e.g., Little, 1997) that the chi-square difference test may be too sensitive. Other statistics are available, including the Akaiki information criterion (AIC; Akaiki, 1987), corrected Akaiki information criterion (Bozdogan, 1987), expected cross-validation index (ECVI; Browne & Cudeck, 1993), and Bayesian information criterion (BIC; Muthen & Muthen, 1998). The values of these indices carry no meaning in and of themselves; they are used to determine the best of a series of models, with preference given to models having values closer to 0. The researcher consults the chi-square difference test results in conjunction with these indices to reach conclusions about which model best fits the data.

What remains before we proceed to our example using real data is a brief discussion of exactly what constraints should be applied and in what order to examine gender differences in factor structures. Equality constraints can be applied to intercepts, factor loadings, residuals, factor variances, and factor covariances. Moreover, it is not necessary to constrain all intercepts, all factor loadings, all residuals, and so on, across genders; rather, particular pairs within each of these categories might be chosen, based on some rationale or theoretical basis. Joreskog and Sorbom (1993b) provide a guide to conducting tests of equivalence using the LISREL software, beginning with an evaluation of the equivalence of the observed variances and covariances across groups, then proceeding to evaluations of the equivalence of the number of factors and pattern of loadings (configural invariance), the values of the loadings (factorial or metric invariance), the values of the residuals, and, finally, the values of the factor variances and covariances. Marsh (1994) proposed a slightly different progression, beginning with the test of configural invariance, then factorial invariance, followed by equivalence of factor covariances, variances, and residuals. He commented that invariance of factor variances and invariance of measurement residuals are usually of lesser substantive importance to the equality of factor structures.

Meredith (1993) and Little (1997) argued that a demonstration of "strong factorial invariance" requires the equality of both intercepts and factor loadings. They specifically noted that the residuals should be free to vary across groups, reasoning that the imposition of equality constraints on the residuals would introduce bias into estimates of other model parameters. They also distinguished between a measurement level of invariance, concerned with intercepts and factor loadings, and a latent level of invariance, concerned with factor means, variances, and covariances. Byrne et al. (1989) introduced partial factorial invariance as a less stringent requirement in exploring the equivalence of factor structures. From their perspective, it might be sufficient to demonstrate that equality constraints on a single intercept and a single factor loading are all that are required. Hancock et al. (1999) supported this position. Obviously, consensus on what constitutes equivalence of factor structures is not yet achieved.

Demonstration of Multiple-Group Confirmatory Factor Analysis Using Existing Data

Description of the Data

For the purposes of example, we now introduce analyses on a portion of data from a longitudinal study of Gulf War veterans. The original data collection occurred within 5 days of the soldiers' return from the Gulf region in 1991 (for additional details, see Wolfe, Erickson, Sharkansky, King, & King, 1999). Follow-up data collections occurred in 1992–1993 and 1997–1998. Data presented here were obtained in the initial 1991 assessment. The sample consisted of 2,949 Army personnel deployed from and returning to Ft. Devens, Massachusetts: 240 women and 2,702 men, with 823 from regular active duty status, 587 from the Reserves, and 1,505 from the National Guard. The focal instrument was the Mississippi Scale for Combat-Related PTSD (Keane et al., 1988), modified for Gulf War personnel (Wolfe, Brown, & Kelley, 1993). To provide a reasonably straightforward demonstration of multiple-group confirmatory factor analysis, we created four composite, or subscale, scores from responses to the 35 items on this scale: a simple average item score across 11 items measuring reexperiencing and situational avoidance, 11 items measuring withdrawal and numbing, 8 items measuring arousal and lack of behavioral or emotional control, and 5 items measuring self-persecution (guilt and suicidality) (King & King, 1994). The primary research question of interest is: Does the structure of PTSD, as measured by the four subscales of the Mississippi Scale, differ as a function of gender? Placed in terms of the guiding theme of this chapter, does the relationship between the hypothetical construct of PTSD and its four manifest indicators differ for women and men?

Analyses

Analyses were accomplished using Arbuckle's (1997) AMOS software package. A series of nested, single-factor means and covariance structures, multiple-group confirmatory factor analyses was conducted using maximum likelihood estimation. The most saturated model, the one with the largest number of estimated parameters, was one in which both the intercepts and loadings, as well as the residuals, were estimated. For the next model in the series, we constrained the factor loadings to be equivalent for women and men. The third model constrained the intercepts to be equivalent. And the fourth model placed equality constraints on the factor variances. Because the model contained only one factor, there were no factor covariances to be considered. Various indices of fit and relative fit, available from the AMOS software, were used to evaluate the strengths of the several models.

Results and Their Interpretation

Table 15.3 presents the results of analyses of the nested, multiple-group models. The first row of Table 15.3 summarizes the findings for the most saturated model and supports an assertion for configural invariance across genders; that is, for both women and men, a single-factor model with four manifest indicators of the PTSD construct appears to provide a reasonably good representation of the observed data. Evidence for this assertion is best seen in the values for the two incremental fit indices, the TLI and the CFI. Both exceed the .95 criterion of good fit recently mandated by Hu and Bentler (1998). Moreover, the value for the RMSEA is less than .05, another standard of good fit set forth by Browne and Cudeck (1993).

The second row of Table 15.3 reports the findings for a model in which the factor loadings are equivalent across genders. When this model is compared to the most saturated model, the chi-square difference ($\Delta\chi^2$) is nonsignificant, indicating that constraining loadings to be equivalent and the attendant change in degrees of freedom (Δdf), do not significantly damage model–data fit. Furthermore, there is a drop in the values of the AIC and the ECVI, the two indices specifically intended to compare models. These decreases also suggest that the more parsimonious, equal-factor loadings model is a better representation of the data. The RMSEA likewise

TABLE 15.3. Results of Multiple-Group Confirmatory Factor Analysis

Model	χ^2	*df*	TLI	CFI	RMSEA	90% CI	AIC	ECVI	$\Delta\chi^2$	*df*
1. Saturated	31.501	4	.996	.999	.048	.034–.065	79.501	.027		
2. Equal factor loadings	35.673	7	.998	.999	.037	.036–.050	77.673	.026	4.172	3
3. Equal factor loadings and equal intercepts	151.561	10	.992	.996	.069	.060–.079	187.561	.064	115.888*	3
4. Equal factor loadings, equal intercepts, and equal factor variances	165.216	11	.992	.996	.069	.060–.079	199.216	.068	13.665*	1

Note. TLI, Tucker–Lewis Index; CFI, comparative fit index; RMSEA, root mean square error of approximation.
*$p < .001$.

decreases in value, a good sign that fit is enhanced with this model. TLI and CFI remain high.

On the other hand, adding the constraint that the intercepts be equivalent across genders is probably not advisable. As shown in the third row of Table 15.3, when this constraint is added, there is a significant increase in the chi-square, accompanied by large increases in the AIC, ECVI, and RMSEA. The RMSEA increases above the .05 recommended cutpoint. In fact, its 90% confidence interval (.060–.079) does not enclose the value .05. Thus, it is better to consider leaving the separate intercepts for women and men free to vary, and conclude that they differ from one another. Information in the fourth row of Table 15.3 indicates that it is also unwise to constrain the variance of the PTSD factor to be equivalent across genders. Consequently, the model that best fits the data is one in which the factor loadings for women and men are equivalent but the intercepts and factor variances are different.

The models of best fit are depicted in Figures 15.3 and 15.4. These displays incorporate what is known as reticular action model (RAM) theory (McArdle & Boker, 1990; McArdle & McDonald, 1984) conventions for presenting a structural equation model. Equations in the form of Equation 11 can be written directly from Figures 15.3 and 15.4. On each display, the large circle represents the PTSD latent variable; the squares represent the four manifest indicators, our observed subscale scores; the smaller circles represent residuals; the small horseshoe-shaped, double-headed arrow, a sling, associated with the PTSD latent variable, represents the factor variance; the same type of horseshoe-shaped, double-headed arrows attached to the residuals represent residual variance; and the triangle is a constant

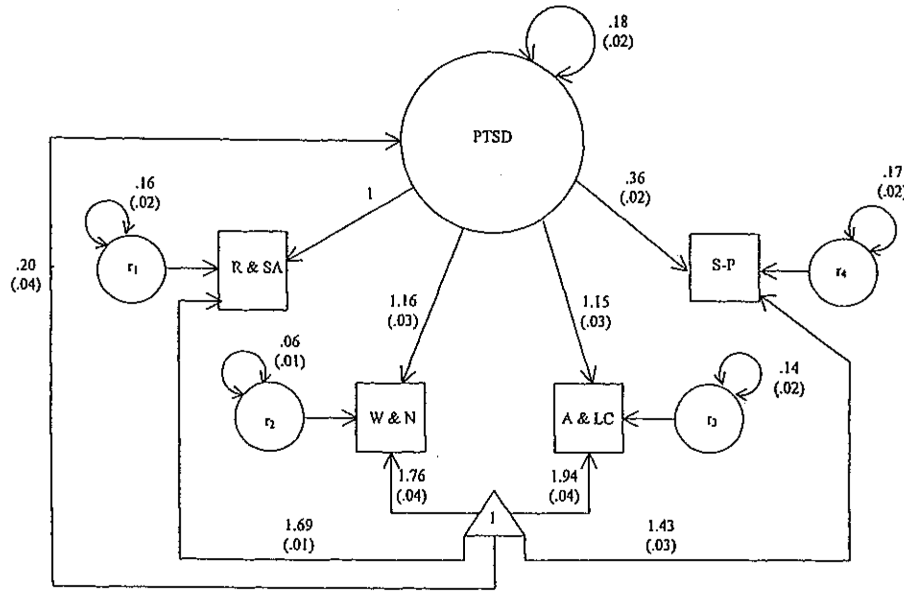

FIGURE 15.3. Graphical representation of the factor-analytic model for women. PTSD, posttraumatic stress disorder; R & SA, reexperiencing and situational avoidance; W & N, withdrawal and numbing; A & LC, arousal and lack of control; S-P, self-persecution (guilt and suicidality).

used to represent means and intercepts. Values are also shown for all parameter estimates, with their standard errors in parentheses. All parameter estimates for both women and men have critical ratios (the estimate divided by its standard error) that exceed 2.00. Although the sampling distribution of these critical ratios is not exactly known, Joreskog and Sorbom (1993b) noted that a value greater than 2.00 suggests statistical significance at $p < .05$.

Note that the path from the PTSD latent variable to the reexperiencing and situational avoidance manifest indicator has a value of 1, with no standard error for either women or men. Note also that the value for the intercept of the regression of PTSD on this manifest indicator is the same for women as for men, and that the mean of the PTSD latent variable for men is equal to 0. These specific constraints set the origin and scale for the latent PTSD variable, a necessity for model identification. The issue of model identification is very complex; the reader is referred to some of the more comprehensive texts on structural equation modeling (e.g., Bollen, 1989; Kline, 1998; Loehlin, 1998; Schumacker & Lomax, 1996). Many of the commercially available software programs set these particular constraints by default.

Figures 15.3 and 15.4 also provide information about the difference between group means on the latent variable. The mean for the latent variable for men is equal to 0 (see preceding paragraph), and the corresponding mean for women is .20. Its associated critical ratio (.20/.04 = 5.00) represents a test of the significance of this estimate from 0. Therefore, for this example, women's PTSD scores exceeded those of men.

As pointed out by Horn and Meredith (1998), an interpretation of the intercepts in a factor-analytic model is that they are a function of the means of the component of the residual that is specific to the manifest variable itself (and not held in common with the latent factor, in our example, PTSD). For purposes of demonstration, we conducted only an overall test of the equivalence of the intercepts (third row, Table 15.3) and found that somewhere among the intercepts, there is at least one difference across genders. We could go further to evaluate the source of the overall difference: for example, if the single intercept for withdrawal and numbing for women (note the 1.76 value in Figure 15.3) is equivalent to that for men (note the 1.91 value in Figure 15.4), or if the intercept for arousal and lack of control for women (1.94) is equivalent to that for men (1.72). Gender differences in

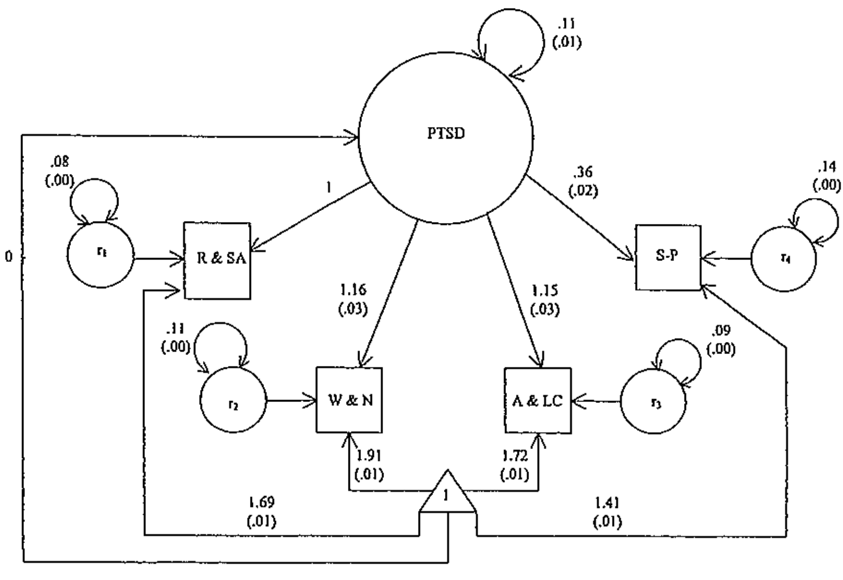

FIGURE 15.4. Graphical representation of the factor-analytic model for men. PTSD, posttraumatic stress disorder; R & SA, reexperiencing and situational avoidance; W & N, withdrawal and numbing; A & LC, arousal and lack of control; S-P, self-persecution (guilt and suicidality).

the intercepts for a particular symptom cluster would suggest that, above and beyond what the four symptom clusters hold in common with one another (presumably, PTSD), women and men differ on another component specific to that cluster, but outside what is PTSD.

Additional Comments and Recommended Resources

Our presentation of multiple-group confirmatory factor analysis might appear somewhat different from most of the published literature that includes this analytic procedure. Typically, confirmatory factor analysis represents what is formally called the analysis of covariance structures, the goal of which is to best reproduce the matrix of variances and covariances among observed scores. We have presented an augmented approach, the analysis of means and covariance structures, the goal of which is to best reproduce the vector of means of the observed scores as well. While slightly more complicated, advantages are obvious in the form of additional information about possible group differences on specific factors, as represented by differences in the intercepts, and group differences in the means of the latent variable(s). Moreover, as suggested by Yung and Bentler (1999), incorporating mean structures into a multiple-group confirmatory factor analysis may yield more stable estimates of the factor loadings.

We also might add that examining differences in relationships between women and men can go beyond confirmatory factor analysis, the measurement component of structural equation modeling, and be applied to the structural component of structural equation modeling. In the latter situation, the concern would be whether associations between latent variables in a theoretical network differ across genders. The basic strategy would be quite similar to that demonstrated in this chapter for the measurement component. Subsequent to establishing factorial invariance, the researcher could test for gender-based differences by means of equality constraints placed on path coefficients and intercepts in the structural model.

As with our discussion of moderated multiple regression, this presentation of multiple-group confirmatory factor analysis is rather cursory. We strongly encourage the interested reader attempting this type of procedure to seek out more extensive treatments of the topic. Among a growing list of very good texts on the larger subject of structural equation modeling are the works of Bollen (1989), Hayduk (1987, 1996), Kline (1998), Loehlin (1998), and Schumacker and Lomax (1996). Additionally, the books edited by Bollen and Long (1993), Hoyle (1995), and Marcoulides and Schumacker (1996) contain chapters that are excellent summaries of special issues in the field. A very approachable ("how to") text on multivariate statistics by Tabachnick and Fidell (1996) contains a chapter on structural equation modeling. McArdle (1996) provided a succinct overview of the confirmatory factor-analytic methodology, directed at the general community of scientists and academics. Finally, Keith (1997) describes an application of confirmatory factor analysis to the process of understanding the construct of intelligence; a similar approach could be taken to explore the structure underlying PTSD and to understand gender differences in PTSD. Journals directly devoted to new developments in structural equation modeling include *Structural Equation Modeling: A Multidisciplinary Journal*, *Multivariate Behavioral Research*, *Applied Psychological Measurement*, and the *Journal of Educational and Behavioral Statistics*. Software available for structural equation modeling includes AMOS (Arbuckle, 1997), EQS (Bentler, 1995), Mplus (Muthen & Muthen, 1998), Mx (Neale, 1993), and LISREL 8/SIMPLIS/PRELIS 2 (Joreskog & Sorbom, 1993a, 1993b, 1993c).

REFERENCES

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Akaiki, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317–332.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*, 411–423.

Arbuckle, J. L. (1997). *AMOS user's guide*. Chicago: SPSS.

Beere, C. A. (1990). *Gender roles: A handbook of tests and measures*. New York: Greenwood Press.

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107*, 238–246.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., & Forneris, C. A. (1996). Psychometric properties of the PTSD Checklist (PCL). *Behaviour Research and Therapy, 34*, 669–673.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Bozdogan, H. (1987). Model selection and Akaiki's information criteria (AIC). *Psychometrika, 52*, 345–370.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466.

Champoux, J. E., & Peters, W. S. (1987). Form, effect size, and power in moderated regression analysis. *Journal of Occupational Psychology, 60,* 243–255.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249–253.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Cohen, J., Cohen, P., Aiken, L. S., & West, S. G. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Darlington, R. B. (1990). *Regression and linear models.* New York: McGraw-Hill.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.) New York: Wiley.

Fitzpatrick, M. F., Salgado, D. M., Suvak, M. K., King, L. A., & King, D. W. (2002). *Associations of gender and gender-role ideology with behavioral and attitudinal features of intimate partner aggression.* Manuscript under review.

Hancock, G. R., Stapleton, L. M., & Berkovits, I. (1999, April). *Loading and intercept invariance within multisample covariance and mean structure models.* Paper presented at the meeting of the American Educational Research Association, Montreal, Canada.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances.* Baltimore: Johns Hopkins University Press.

Hayduk, L. A. (1996). *LISREL: Issues, debates, and strategies.* Baltimore: Johns Hopkins University Press.

Horn, J., & Meredith, W. (1998, October). *Factorial invariance in longitudinal research.* Paper presented at the conference on New Methods for the Analysis of Change, State College, PA.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Hoyle, R. H. (Ed.). (1995). *Structural equation modeling: Concepts, issues, and applications.* Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P.M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model specification. *Psychological Methods, 3,* 424–453.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Thousand Oaks, CA: Sage.

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin, 117,* 348–357.

Joreskog, K. G., & Sorbom, D. (1993a). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Chicago: Scientific Software.

Joreskog, K. G., & Sorbom, D. (1993b). *LISREL 8 user's reference guide.* Chicago: Scientific Software.

Joreskog, K. G., & Sorbom, D. (1993c). *PRELIS 2 user's reference guide.* Chicago: Scientific Software.

Keane, T. M., Caddell, J. M., & Taylor, K. L. (1988). Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology, 56,* 95–102.

Keith, T. Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 373–402). New York: Guilford Press.

King, D. W., King, L. A., Foy, D. W., & Gudanowski, D. M. (1996). Prewar factors in combat-related posttraumatic stress disorder: Structural equation modeling with a national sample of female and male Vietnam veterans. *Journal of Consulting and Clinical Psychology, 64*; 520–531.

King, D. W., King, L. A., Gudanowski, D. M., & Vreven, D. L. (1995). Alternative representations of war zone stressors: Relationships to posttraumatic stress disorder in male and female Vietnam veterans. *Journal of Abnormal Psychology, 104*, 184–196.

King, L. A., & King, D. W. (1993). *Manual for the Sex-Role Egalitarianism Scale: An instrument to measure attitudes toward gender-role equality*. London, Ontario: Research Psychologists Press/Sigma Assessment Systems.

King, L. A., & King, D. W. (1994). Latent structure of the Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: Exploratory and higher-order confirmatory factor analyses. *Assessment, 1*, 275–291.

King, L. A., & King, D. W. (1997). Sex-Role Egalitarianism Scale: Development, psychometric properties, and recommendations for future research. *Psychology of Women Quarterly, 21*, 71–87.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Kulka, R. A., Schlenger, W. F., Fairbank, J. A., Hough, R. L., Jordan, B. K., Marmar, C. S., & Weiss, D. S. (1990). *Trauma and the Vietnam War generation: Report on the findings from the National Vietnam Veterans Readjustment Study*. New York: Brunner/Mazel.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53–76.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analyses* (3rd ed.). Mahwah, NJ: Erlbaum.

Lunneborg, C. E., & Abbott, R. D. (1983). *Elementary multivariate analysis for the behavioral sciences: Applications of basic structure*. New York: Elsevier/North-Holland.

Marcoulides, G. A., & Schumacker, R. E. (Eds.). (1996). *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Erlbaum.

Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling: A Multidisciplinary Journal, 1*, 5–34.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391–410.

McArdle, J. J. (1996). Current directions in structural factor analysis. *Current Directions in Psychological Science, 5*, 11–18.

McArdle, J. J., & Boker, S. M. (1990). *RAMpath: Path diagram software*. Denver: Data Transforms.

McArdle, J. J., & Cattell, R. B. (1994). Structural equation models of factorial invariance in parallel proportional profiles and oblique confactor problems. *Multivariate Behavioral Research, 29,* 63–113.

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology, 37,* 234–251.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376–390.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543.

Mulaik, S. A., & James, L. R. (1995). Objectivity and reasoning in science and structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Muthen, L. K., & Muthen, B. O. (1998). *Mplus user's guide.* Los Angeles: Author.

Neale, M. C. (1993). *Mx: Statistical modeling.* Richmond: Medical College of Virginia.

O'Connor, B. P. (1998). SIMPLE: All-in-one programs for exploring interactions in moderated multiple regression. *Educational and Psychological Measurement, 58,* 836–840.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Fort Worth: Harcourt Brace.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling.* Mahwah, NJ: Erlbaum.

Schumacker, R. E., & Marcoulides, G. A. (Eds.). (1998). *Interaction and nonlinear effects in structural equation modeling.* Mahwah, NJ: Erlbaum.

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family, 38,* 15–28.

Spence, J. T., & Helmreich, R. (1972). The Attitudes Toward Women Scale: An objective instrument to measure the attitudes toward the rights and roles of women in contemporary society [Ms. No. 153]. *JSAS: Catalog of Selected Documents in Psychology,* 2, 66–67.

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family, 41,* 75–88.

Steiger, J. H. (1990). Structural model evaluation and modification: An internal estimation approach. *Multivariate Behavioral Research, 25,* 173–180.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50,* 253–264.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Weathers, F. W., Keane, T. M., King, L. A., & King, D. W. (1997). Psychometric theory in the development of posttraumatic stress disorder assessment tools. In J. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD* (pp. 98–135). New York: Guilford Press.

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). *The PTSD Checklist: Reliability, validity, and diagnostic utility.* Paper presented at the annual meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.

West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality, 64,* 1–48.

Wolfe, J., Brown, P. J., & Kelley, J. M. (1993). Reassessing war stress: Exposure and the Persian Gulf War. *Journal of Social Issues, 49,* 15–31.

Wolfe, J., Erickson, D. J., Sharkansky, E. J., King, D. W., & King, L. A. (1999). Course and predictors of posttraumatic stress disorder among Gulf War veterans: A prospective analysis. *Journal of Consulting and Clinical Psychology, 67,* 520–528.

Yung, Y., & Bentler, P. M. (1999). On added information for ML factor analysis with mean and covariance structures. *Journal of Educational and Behavioral Statistics, 24,* 1–20.

Zedeck, A. (1971). Problems with the use of "moderator" variables. *Psychological Bulletin, 76,* 295–310.